Specialty Conference

N-of-1 Clinical Trials

A Technique for Improving Medical Therapeutics

Discussant
ERIC B. LARSON, MD, MPH, *Seattle*

This discussion was selected from the weekly Grand Rounds in the Department of Medicine, University of Washington School of Medicine, Seattle. Taken from a transcription, it has been edited by Paul G. Ramsey, MD, Associate Professor of Medicine, and Philip J. Fialkow, MD, Professor and Chair of the Department of Medicine.

ERIC B. LARSON, MD*: Most clinicians have a keen interest in therapeutics and especially therapeutic efficacy. In fact, medical therapeutics can be viewed as a series of therapeutic experiments as follows:

A				*B*
Initial State	→	Therapy	→	Subsequent State

The patient comes to the physician in an initial state, *A*, and is offered treatment. The patient then assumes a subsequent state, *B*.[1] If *B* is more desirable, we typically judge that therapy was effective. If *B* is no different or is less desirable, we judge that therapy made no difference or was ineffective. Although this account seems straightforward, such simple assertions may not be true because of confounding factors.[2]

Effectiveness may be overestimated because of several factors. First, a patient can recover spontaneously coincident with treatment, an especially well-known occurrence for self-limited conditions. Second, patients commonly present when their symptoms are worse, especially patients with a chronic disease. Coincidental treatment appears to cause the problem to subside when the patient has simply returned spontaneously to the average, so-called baseline state of a chronic disease. This has been referred to as "regression toward the mean."[3] A third factor that may lead to an overestimation of effectiveness is a placebo effect. For some therapies, as much as 30% or more of the benefits may be due to the well-known placebo effect.[4] Finally, the expectation of a beneficial response and a willingness-to-please effect[5] are related to the placebo effect. In many patients, the simple "expectation" that a treatment will be beneficial may often be sufficient to promote a beneficial effect. The willingness-to-please effect results from the so-called obsequiousness bias[5] in which a patient gets better to please an expectant physician.

Similar confounding forces can obscure therapeutic effectiveness. Coexistent illness can coincidentally exacerbate the underlying problem. Chronic diseases have spontaneous exacerbations, and when these occur coincident with treatment, it appears that therapy is ineffective. Malingering or a secondary gain in which the patient experiences benefit from not getting better can make a patient resistant to the true effect of treatment. An age-related (physiologic) decline superimposed on a beneficial treatment effect may combine to cancel each other. Finally, if an incorrect diagnosis has been made, treatment will appear to be ineffective. For example, if a patient's symptoms or signs represent the upper or lower limits of a normal variation, then the treatment received, although usually effective, is ineffective in the misdiagnosed case.

Randomized Clinical Trials

Fortunately, randomized clinical trials (RCTs) have been used to evaluate medical therapeutics since the late 1940s.[6] Because such trials help eliminate the confounding factors outlined above, they have become the gold standard by which clinicians judge therapeutic efficacy. An RCT allocates consecutive patients to different treatments or randomly allocates the order of treatment in crossover experiments. When done carefully with enough patients, the randomization eliminates bias that might confuse the interpretation of the therapeutic experiment.

Unfortunately, many of a clinician's day-to-day treatment decisions cannot be based on the results of randomized trials. Table 1 shows examples of situations or problems in which RCTs may not be appropriate for making therapeutic choices. Unavailability of randomized clinical trials may be encountered in the case of a rare or unusual disease. Randomized trials may also not be available for some older treatments and for newer or novel treatments. Because RCTs have been widespread only since 1970, older treatments were often not evaluated by them. Newer or novel treatments, especially those devised by clinicians for single patients, are typically not subjected to randomized trials.

Even when there are good randomized trials showing efficacy, several factors limit their generalizability to a specific patient. For example, the patient might be outside the eligibility requirements for entry into an RCT. Eligibility criteria for most trials are so restrictive that less than 10% of patients with the disease in question may be accepted. Not surprisingly, the patients who are excluded are the ones in whom therapeutic dilemmas and an evaluation of therapeutics are often the most troublesome. Thus, their omission

*Professor of Medicine, Department of Medicine, and Medical Director, University of Washington School of Medicine, Seattle.

(Larson EB: N-of-1 clinical trials—A technique for improving medical therapeutics [Specialty Conference]. West J Med 1990 Jan; 152:52-56)

This work is supported by grant 1 RO3 HS 6006 from the National Center for Health Services Research and Health Care Technology Assessment.
Reprint requests to Department of Medicine, RG-20, University of Washington School of Medicine, Seattle, WA 98195.

TABLE 1.—*Limits of Randomized Clinical Trials (RCTs) for Care of Individual Patients*

RCT unavailable or impossible
Good RCTs show benefit but may not be generalizable
Eligibility criteria too restrictive
Some patients are nonresponders
Side effects
Good RCTs show *no* benefit but may not be generalizable
Atypical patients
Treatment response is idiosyncratic

from RCTs allows investigators to assess efficacy with fewer complicating factors. Another problem arises from the fact that even though a randomized trial has shown efficacy, not all patients will benefit from treatment. In addition, some patients may experience enough side effects that the net effect of treatment is harmful. The single patient who does not have a beneficial response experiences that event with 100% certainty even when generalizations based on populations studied by RCTs indicate the net effects are likely to be beneficial.

There are also limits to the generalizability of RCTs that show no apparent benefit. Good randomized clinical trials may not show any net benefit, but an individual patient may still benefit from treatment, especially if the treatment has biologic plausibility. Some RCTs have inadequate sample sizes and, hence, inadequate statistical power to show efficacy.[7] An individual patient could also be an atypical responder, or responsiveness to treatment may be idiosyncratic and difficult to demonstrate by an RCT.

In summary, even though randomized clinical trials are widely used for assessing therapeutic efficacy, their results may not apply to single patients or they may be unavailable for certain treatments, thus leaving clinicians in a quandary about therapeutic efficacy. Because of this quandary, there is increasing interest in single-patient experiments. A number of terms have been used to describe single-patient experiments, including N-of-1 trials, single-patient clinical trials, single-case analysis, crossover and self-controlled research designs, and single-patient RCTs. The field has an interesting history and holds great promise for improving the science of medical therapeutics.

Case Reports

Because case reports can be useful ways to illustrate valuable clinical lessons, I will present three single-case analyses in the order of my exposure to them. The first, a "case report" presented at the American Federation of Clinical Research meetings in 1985, was the case that piqued my interest in single-patient trials.[2] The second, a classic case that occurred at the interface of the developing science of statistics and popular culture, is intriguing for both its contents and the statistical power of its design.[8] The final case illustrates a single-case clinical trial that, although not random and only "single blinded," was convincing and influential.[9]

The first case was reported by Guyatt and co-workers from McMaster University, Hamilton, Ontario.[2] The patient, a 65-year-old man with uncontrolled asthmatic bronchitis, was becoming progressively more disabled by dyspnea with even simple daily activities. His therapeutic regimen eventually consisted of albuterol inhaler, ipratropium bromide, theophylline, and daily doses of prednisone. The clinician and the patient were uncertain whether the theophylline or ipratropium therapy was beneficial. Both suspected that theophylline was helpful and ipratropium was not. To optimize the therapeutic regimen, a single-patient trial was designed. Either theophylline or placebo, in a random order, was given for ten-day crossover periods. Three 10-day crossover pairs were planned. The end points included dyspnea, the need for albuterol inhaler, and the amount of sleep disturbance. During the first period, the patient did better than during the second ten days of the crossover trial. The same pattern then appeared during the second crossover period. The trial, which was originally scheduled to go for three crossover periods (about 60 days), now seemed too long to both the clinician and the patient. Both agreed that the trial should be terminated, presumably to allow the patient to resume taking theophylline. They were surprised when the placebo was associated with scores indicating increased well-being. Based on a review of the literature and the patient's course, it was determined that the seemingly anomalous results were most likely explained by gastroesophageal reflux (a xanthine side effect) and aspiration.[10] The theophylline therapy was stopped, and subsequently an N-of-1 trial of ipratropium revealed the beneficial therapeutic effects of its use. Eventually the patient was treated with a regimen of albuterol and ipratropium. He then tolerated a prednisone taper so that he could comfortably complete most of his activities of daily living on a regimen of 10 mg of prednisone every other day.

The second "case report" is not a medical case but represents a particularly famous single-case experiment. The case was an important one in the development of principles of experimentation and illustrates some useful points about randomization and statistical power. In 1935, R. A. Fisher, a British statistician whose name is most often linked with multiple-subject experiments, reported an example of how to conduct an experiment with a single subject and used that example to explain basic notions that underlie all experiments. This was the "lady tasting tea experiment."[8]

The case involved a tea-drinking English woman who claimed that she could tell whether the tea was added to the milk or the milk was added to the tea. Four cups of tea were prepared one way and four cups the other way, and the eight cups were then presented to her in a random sequence. She was told in advance that she was to identify the four cups that were prepared each way. The lady correctly identified all eight cups, and the P value was determined by the randomization test procedure. The null hypothesis was that her response at any treatment time was the same as it would have been at that time if any of the other cups had been presented. There are $8!/4!4! = 70$ ways in which eight cups can be presented with respect to milk first or tea first, given that four cups were milk first and four tea first. Thus, Fisher computed the P value as 1/70 because only 1 of the possible sequences of 4 *M*s and 4 *T*s correctly matched the woman's responses ($P = .014$).

An important feature of this experiment, in contrast to the first case report, is that the randomization occurred in blocks of eight treatments, not blocks of two as in the typical crossover experiment. Thus, the statistical power was considerably greater.

The third case report is a more primitive example of a single-patient trial.[9] Nonetheless, it also shows the value of single-patient experimentation. The report entitled "Internal-Mammary-Artery Ligation for Coronary Insufficiency—An Evaluation" was based on a presentation made in 1957 to the New England Surgical Society. This topic would later be investigated in a widely quoted article from the University of Washington describing a randomized, single-blind trial that compared a sham operation with internal mammary ligation.[11] Ralph Adams, MD, in the 1958 paper,[9] reported four cases, one of which was of a 60-year-old man admitted "three days after occurrence of his known episode of coronary thrombosis."

> His case was well known to the hospital because of previous attacks of deep thrombophlebitis, pulmonary embolism and hypercholesterolemia, and prior episodes of coronary occlusion. Precordial pain was intense and he was apprehensive that he would die. He was a highly educated man, well informed for a layman, on medical matters and in a position of considerable community responsibility. Admission was for the specific purpose of altering internal mammary circulation in the hope of giving him some cardiac protection. He was told . . . that this procedure was currently being widely discussed and, in some quarters, enthusiastically recommended. He was also informed that the hospital was in the process of evaluating the procedure as definitely as possible. These background facts led him to request that the operation be tried in the hope that he might be helped. . . .
>
> At operation, on the day of admission, a short incision was made in the second intercostal space lateral to each sternal border and each internal mammary artery was exposed. A silk ligature was placed about each artery but neither was tied. Thus, only a first-stage operation had been done, consisting of a skin incision and encirclement but not ligation of the internal mammary arteries.
>
> On awakening from the brief and light anesthetic, the patient reported that he was free of pain. He has had no pain since that date. An electrocardiogram on the day after operation showed no detectable change from preoperative tracing. Two days after the operation the ligatures from the internal mammary arteries were tied. Subsequent electrocardiographic tracings gave no evidence of improvement.

The author goes on to describe follow-up, which included no recurrence of symptoms, and states that

> in this case, there was not a fair chance to assay the relief of symptoms to be obtained by internal mammary artery ligation because the patient lost all symptoms after the first portion of a staged procedure that he believed to be the completed operation.

Adams reported what we would call a nonrandomized single-patient crossover experiment. A sham operation was followed by a real operation—dramatically showing what many might now call a placebo effect of internal mammary exposure.

Formation of an N-of-1 Clinical Trial Service

Before establishing a single-patient trial service, we contacted Dr Gordon Guyatt, who has actively investigated single-patient trials. He provided us with great encouragement and a summary of the experience of an N-of-1-trials service at McMaster University.[2] Most of his trials had been in the subspecialties of pulmonary medicine and rheumatology. Of the first 42 trials done at the center, 29 gave definitive results. In 11, active treatment was found to be effective, in 17 it was ineffective, and in 1 it was harmful (the theophylline case). Eight other trials gave less definitive results. Five were judged unsuccessful, three because, despite definitive outcomes, the results did not lead to action (G. Guyatt, written communication, June 1987).

Based on this encouraging report, we submitted a small grant proposal to the National Center for Health Services Research. Our research group, which includes Allan Ellsworth, PharmD; Jim Nuovo, MD (family medicine); Ina Oppliger, MD (rheumatology); Gerald van Belle, PhD; and Alice Arnold, MS (biostatistics), is now funded to establish and evaluate a single-patient trial service. We have announced our intentions to workers in other specialties and are currently receiving patients.

Because the objective of the "N of 1" experiment is to find the best treatment for a particular patient, we and others believe that some of the ethical questions asked of the standard randomized trial no longer apply.[2] For example, does the potential benefit to other patients outweigh the possible risk to this patient? Nonetheless, three ethical requirements do apply. First, a patient's free and informed consent should be requested after the clinician has described every feature of the trial that would materially affect the patient's decision to take part, including the reported effectiveness and safety of alternative treatments, the treatment targets to be used, and the duration and number of treatment periods to be executed. The second ethical requirement is that a patient must be free to withdraw at any time without loss of care. The third is that the same degree of confidentiality applied in other clinical situations must apply to the study results. One of our first tasks as an N-of-1 clinical trial service was to approach the Human Subjects Committee (Institutional Review Board) and seek approval for pending single-patient trials. They have developed an expedited approval process that facilitates the prompt institution of clinical trials.

When to Do a Clinical Trial

Perhaps the most germane issue in single-patient trials is when to do them. That is, when is a patient most likely to benefit from the results of a single-patient trial? The most important issue here is whether there is doubt about efficacy. Doubt may occur because neither the patient nor the physician is certain an existing treatment is working. In this setting, a patient with a chronic disease may be doing poorly or not improving on a medication regimen that could also be causing side effects, as exemplified by the theophylline case.

Another instance when efficacy may be in doubt is during the institution of a new treatment. Here the patient is being offered a new drug and the question is, "Will it work?" The clinician may be uncertain when the literature is equivocal about the drug, the risk-to-benefit ratio is less favorable, or the patient is reluctant to comply with presumably efficacious treatment.

For patients with rare or unusual conditions, the use of the single-patient trial may not only benefit the patient but also add to knowledge about the management of unusual conditions. The literature contains numerous examples of single-patient experiments where treatments of conditions like familial Mediterranean fever and narcolepsy were evaluated with N-of-1 trials.

Doubt about efficacy may be a motivating factor for a single-patient trial also when a patient insists on a treatment as necessary or effective in contradiction to medical advice or practice. The single-patient trial can be used when the physician is unable to convince the patient otherwise. In this case, a negative clinical trial should not surprise the physician but may be convincing to the patient.

After determining whether therapeutic efficacy is in doubt and deciding whether one wishes to demonstrate efficacy or a lack thereof, the clinician will need to consider other questions that affect the feasibility and worth of a single-patient trial. First is whether a treatment will likely be long term. Given the time required to conduct such a trial, single-patient trials of short-term therapies tend not to be worth the effort required of the patient, and they are less likely to have value for the individual patient unless the patient will require the short-term treatment repeatedly.

Several questions related to the pharmacokinetics of a possible therapeutic agent affect the logistics and ease of doing single-patient trials.[12] The ideal treatment for single-patient trials is one that can be rapidly started and stopped. Thus, outcomes can be assessed starting relatively early in the trial, and there is little or no carryover between treatment periods. When these criteria are not met, carryover or period effects may complicate the interpretation.[12] These effects may require trials that are much more time consuming (for example, involving washout periods) or involve special design modifications. In general, single-patient trials are less likely to be useful for curative treatments (so-called period effects) or for long-acting treatments (due to carryover effects).

How to Do a Clinical Trial

There are three critical components of the single-patient trial: randomization, blinding of patient and physician to treatment assignment, and defining and quantitating the outcomes. The last, establishing explicit criteria for evaluating the efficacy of treatment, is a feature of the single-patient trial that is also important for medical therapeutics in general.

Randomization is necessary to minimize systematic biases that will occur related to the order of treatment and to permit double blinding to occur. Randomization is usually accomplished in a crossover style, that is, in blocks of two. If, however, it is predetermined that four, six, or eight trials will be done, the statistical power of the trial is improved considerably by randomization in larger blocks.[13] For example, when six trials are planned, the possible P values range from .125 for the paired experiment in which three crossover pairs occur ($[1/2]^3$) to .03 when all six trials are randomized independently ($[1/2]^6$). Intermediate values are possible when constraints are added.

Blinding is a key element to minimize observer-induced bias. In most single-patient trials, the patient records symptoms and, in some cases, signs. Ideally both patient and physician are blind to the treatment assignment. Records of assignment are kept with one of the trial service staff and, if a drug is involved, the pharmacist who has prepared the treatment packages.

Single-patient trials require that the goals of treatment be explicitly identified at the time the patient enters the trial. Ideally, three to five key variables are determined. The variables may reflect disease activity or symptom severity. Usually the most important variables measure patient functioning, reflecting the value of treatment for the patient. In the ideal case, outcomes would include the measurement of a physical sign, a subjective or objective rating of performance in conjunction with, for example, a laboratory measurement reflecting disease activity. The patient's goals must be assayed to be certain that the measures of performance are compatible with the patient's wishes, especially regarding quality of life.

Systematic measurement of a limited number of variables is important for a successful single-patient trial. We typically use self-administered questionnaires that rely on 7-point Likert scales or tabulate the frequency of events. We also teach patients to measure biologic variables like the forced expiratory volume in one second, peak flow, and walk time. We have found it easier to use 7-point Likert scales than visual analog scales. In the standard crossover design, the patient can be asked to state a preference for one treatment period compared with the other.

There are other issues that must be solved when designing a clinical trial. A critical question is the duration of treatment. In general, we believe the old adage, "shortest is easiest." Treatment often takes longer than expected, however, because time is required for peak effects to develop or for treatment effects to dissipate. For drug regimens that are rapidly started and stopped, treatments can be shorter and a random block design of six or eight trials of active drug and placebo can be evaluated in less than two weeks.

A special case occurs when a drug is being used to minimize or prevent attacks or exacerbations of a recurrent disease. To determine duration, the frequency of exacerbation needs to be estimated. Given a reasonable estimate of the frequency, the duration can be based on the "rule of 3s." This rule states that if an event occurs once every x days, the duration of observation must be three times x days to be 95% certain to observe one event. In the case of familial Mediterranean fever where an attack may occur once every two weeks, the treatment period would need to last six weeks to be reasonably certain to observe an effect.

Another question that affects the duration of the trial is how many pairs or trials are needed. The answer to this is the tautology, "as many as are needed." In some trials, we have recommended that a single pair may provide an adequate demonstration of efficacy. Such a demonstration lacks statistical power, but the demonstration of effect may be so compelling as to convince both patient and physician that efficacy is no longer in doubt. On the other hand, when the probability of a treatment being effective is about 50% before the

TABLE 2.—*Posterior Probabilities as Function of Prior Probabilities and Likelihood Ratio*

Prior Belief Treatment Is Effective, P	*Likelihood That Treatment Is Better Than Spontaneous*	*Patient Improves*	*Posterior Probability, P*
.01	3	Yes	.030
	5	Yes	.051
	1/3	No	.003
	1/5	No	.002
.10	3	Yes	.25
	5	Yes	.55
	1/3	No	.032
	1/5	No	.022
.50	3	Yes	.75
	5	Yes	.83
	1/3	No	.25
	1/5	No	.17
.80	3	Yes	.92
	5	Yes	.95
	1/3	No	.57
	1/5	No	.44
.90	3	Yes	.96
	5	Yes	.98
	1/3	No	.75
	1/5	No	.64
.95	3	Yes	.98
	5	Yes	.99
	1/3	No	.86
	1/5	No	.79

trial, and there are major risks of side effects, anything short of a statistical certainty may not be satisfactory. In the case of a paired crossover trial, the binomial distribution suggests that after four trials, the probability of treatment being repeatedly favored over placebo is .5 after the first trial, .25 after the second trial, .125 after the third trial, and .0625 after the fourth trial, which is $(1/2)^4$.

In general, the issue of "statistical" certainty—the mythical $P < .05$—is less critical in single-patient trials. An interesting perspective is added by assaying the clinician's estimate of the likelihood of success in that patient (the prior probability) and determining the estimated likelihood that the treatment is efficacious based on the literature. Using a Bayesian analysis, a posterior probability based on the patient outcome in a single-patient trial can be calculated as shown in Table 2 (G. van Belle, written communication, June 1987). These posterior probabilities show the effect that a single-patient trial can have on a clinician's level of certainty that treatment will be helpful for a patient.

Conclusion

We formed the trial service to simultaneously establish, demonstrate, and determine the value of single-patient trials in clinical practice and to help do the clinical trials. Our involvement ranges from being limited consultants providing study drugs and simply reviewing the protocol, to providing detailed, in-depth consultation regarding the value of a clinical trial in a particular patient, developing a study design, interviewing the patient, developing target outcomes, printing forms, preparing placebo drug and outcome forms, and doing follow-up. In all cases, we provide an interpretation of the results of the trial and are anxious to learn how the trial was used in clinical decision making and practice.

In summary, single-patient clinical trials can be used to improve the efficacy of treatment—especially long-term treatments and treatments with uncertain efficacy or a risk of serious toxic effects. Examples of suitable conditions for study are numerous, including common problems such as chronic obstructive lung disease, osteoarthritis, recurrent headache and other chronic pain syndromes, "fibrositis" or fibromyalgia, and agitation in demented patients. We have done trials in these common conditions and have also investigated more unusual and complex problems such as progestational drug side effects, treatment of the "restless" leg syndrome, and treatments of orthostatic hypotension. The principal benefits are an increased certainty for patients and their physicians that a treatment is worth pursuing because it is effective or should be abandoned because of an absence of a net benefit.

REFERENCES

1. Feinstein AR: Clinical biostatistics—II. Statistics versus science in the design of experiments. Clin Pharmacol Ther 1970; 11:282-292
2. Guyatt G, Sackett D, Taylor DW, et al: Determining optimal therapy—Randomized trials in individual patients. N Engl J Med 1986; 314:889-892
3. Sackett DL: Clinical diagnosis and the clinical laboratory. Clin Invest Med 1978; 1:37-43
4. Brody H: The lie that heals: The ethics of giving placebos. Ann Intern Med 1982; 97:112-118
5. Sackett DL: Bias in analytic research. J Chronic Dis 1979; 32:51-63
6. Cochrane AL: Effectiveness and Efficiency: Random Reflections on Health Service. London, Nuffield Hospitals Trust, 1972
7. Freiman JA, Chalmers TC, Smith H, et al: The importance of beta, the type II error, and sample size in the design and interpretation of clinical trials. N Engl J Med 1978; 299:690-695
8. Edgington ES: Statistics and single case analysis. Prog Behav Modif 1984; 16:83-119
9. Adams R: Internal-mammary-artery ligation for coronary insufficiency. N Engl J Med 1958; 258:113-116
10. Berquist WE, Rachelefsky GS, Kadden M, et al: Effect of theophylline on gastroesophageal reflux in normal adults. J Allergy Clin Immunol 1981; 67:407-411
11. Cobb LA, Thomas GI, Dillard DH, et al: An evaluation of the internal-mammary-artery ligation by a double-blind technique. N Engl J Med 1959; 260:1115-1118
12. Porta MS: The search for more clinically meaningful research designs: Single-patient randomized clinical trials. J Gen Intern Med 1986; 1:418-419
13. Kazdin AE: Single-Case Research Designs: Methods for Clinical and Applied Settings. New York, Oxford Press, 1982